The importance of data

Data is the heart of GIS. How it is created, stored and shared are endlessly debated, because these can have consequences on the correctness, timeliness and relevance of analyses results.

Who creates it?

Generally, it is created by private companies and governments. Private data is harder to access, unless it is shared explicitly. Local, regional and state governments do create and share a lot of data, which generally can be used by analysts free of charge. They don't always make it easy to download the data. Often, they publish web maps that have no explicit download option.

How can we get it for our own uses?

For analysis, it is generally still more useful to download data locally and either use an Esri product or a spatial database a la PostGIS to store and analyze data. To get this data, one can use open data portals, but often available data is only published as web services, which can be queried using URL endpoints and parameters. To get around these limitations, we can use Python.

How data is shared and used

Most data is shared as files, but with the increasing proliferation of web maps, lots of data is shared as web services, published using ArcServer/ArcEnterprise or GeoServer, an open source alternative. As most local governments use Esri products, we will focus today on getting data from ArcServer web map services. However, we do not need an ArcGIS license to download this data!

Files vs services

Generally analysis is performed on files or in database tables, on a local system or computer. The main problem with data is keeping it up to date, and files are terrible as they are generally a copy of the main data source, a “dead branch” cut off the main tree. Databases are a little better at ensuring that data edits are stored as a history, with the ability to reverse engineer edits and to also provide ease of analysis, but they are hard to share with large groups of users, especially external users.

Services make it easy to 'publish' data stored in files or databases. Publishing data services means sharing it on the internet with a URL to display and query in a web map or a GIS software such as ArcMap or QGIS. URL Parameters are used to specify the exact data desired and are passed to the database, which is queried in SQL. The data is returned as JSON or as images.

Open data portals

Open data portals are nice. They are often the best and most well organized place to find data from governments, with a few caveats: the data available is often not up to date, and not all data published by a government are not available all the time.

What if the data isn't shared on a portal?

Then we have to get it ourselves, if it's published as a data source. That's where Python can help!

Map Service Reader:

https://github.com/geolibrerian/mapservicereader

This is the code for the talk. It should be downloaded and installed in the folder that you will use for this process.

Required Modules:

PySHP - https://github.com/GeospatialPython/pyshp

This library allows for the reading, writing and editing of shapefiles, without the need for ArcGIS licences.

Note: this module has just been updated; use the older version 1.2.11 which is included in the repo already

To install:

in a terminal (Linux/MacOS) or command line, run “pip install pyshp”

Requests- https://github.com/requests/requests

The useful requests library improves on the native Python standard library internet modules (urllib and httplib). It's so useful, it has been made a part of the standard library for Python 3.

For *Python 2* users, run “pip install requests” to install the module.

The modules can also be installed using the setup.py file included with the modules (pip automates this process) if pip can't be located.